

ORIGINAL CONTRIBUTION

A Silicon Model of Early Visual Processing

CARVER A. MEAD AND M. A. MAHOWALD

California Institute of Technology

(Received and accepted 5 October 1987)

Abstract—An analog model of the first stages of retinal processing has been constructed on a single silicon chip. Each photoreceptor computes the logarithm of the incident light intensity. A resistive network is used to compute a spatially smoothed version of the receptor outputs. An amplified difference between the receptor signals and their smoothed counterparts forms a second-order spatial filter. Measured outputs from an experimental 48×48 pixel array show many of the characteristics of the bipolar cells in vertebrate retina.

Keywords—Retina, Machine vision, Analog CMOS, Spatial filter, Neural model.

INTRODUCTION

Many of the most striking phenomena known from perceptual psychology are a direct result of the first levels of neural processing. In the visual systems of higher animals, the well-known *center-surround* response to local stimuli is responsible for some of the strongest visual illusions. For example, Mach bands, the Hermann-Hering grid illusion, and the Craik-O'Brian-Cornsweet illusion can all be traced to simple inhibitory interactions between elements of the retina (Ratliff, 1965). The high degree to which a perceived image is independent of the absolute illumination level can be viewed as a property of the mechanism by which incident light is transduced into an electrical signal. We present a model of the first stages of retinal processing in which these phenomena are viewed as natural by-products of the mechanism by which the system adapts to a wide range of viewing conditions. Our retinal model is implemented as a single silicon chip, which contains integrated photoreceptors and processing elements; this chip generates, in real time, outputs that correspond directly to signals observed in the corresponding levels of biological retinas.

RETINAL STRUCTURE

Because our model of retinal processing is implemented on a physical substrate, it has a straightforward structural relationship to the retinas of higher animals. A thorough review of the biological literature up to 1973 can be found in *The Vertebrate Retina* (Rodieck, 1973), and more recent work in *The Retina: An Ap-*

proachable Part of the Brain (Dowling, 1987). Although each animal is unique in detail, the gross structure of the retina has been conserved throughout the vertebrates.

The major divisions of the retina can be seen in the cross section shown in Figure 1. Light is transduced into an electrical potential by the photoreceptors at the top. The primary signal pathway proceeds from the receptors through the triad synapses to the invaginating bipolar cells, and thence to the ganglion cells, the axons of which form the optic nerve. This pathway penetrates two dense layers of neural processes and associated synapses: the outer plexiform layer just below the photoreceptors, and the inner plexiform layer just above the ganglion cell bodies. The horizontal cells are located within the outer plexiform layer, and the inner plexiform layer contains amacrine cells. The horizontal and amacrine cells thus spread across a large area of the retina, in layers transverse to the signal flow. Information in the retina is represented by smoothly-varying analog signals until it reaches the ganglion cell axons where it is encoded in nerve pulses which are quasi-digital (digital in amplitude but analog in time).

Our model is concerned with the processing that occurs in the receptors and the outer plexiform layer. The key processing element in this region is the triad synapse, which is found in the base of the photoreceptor. This synapse is the point of contact between the photoreceptor, the horizontal cells, and the bipolar cells. The computation performed by the model can be stated very simply in terms of these three elements: The photoreceptor takes the logarithm of the intensity. The photoreceptor output is spatially and temporally averaged by the horizontal cells. The bipolar cells' output is proportional to the difference between the photodetector signal and the horizontal cell signal. We will

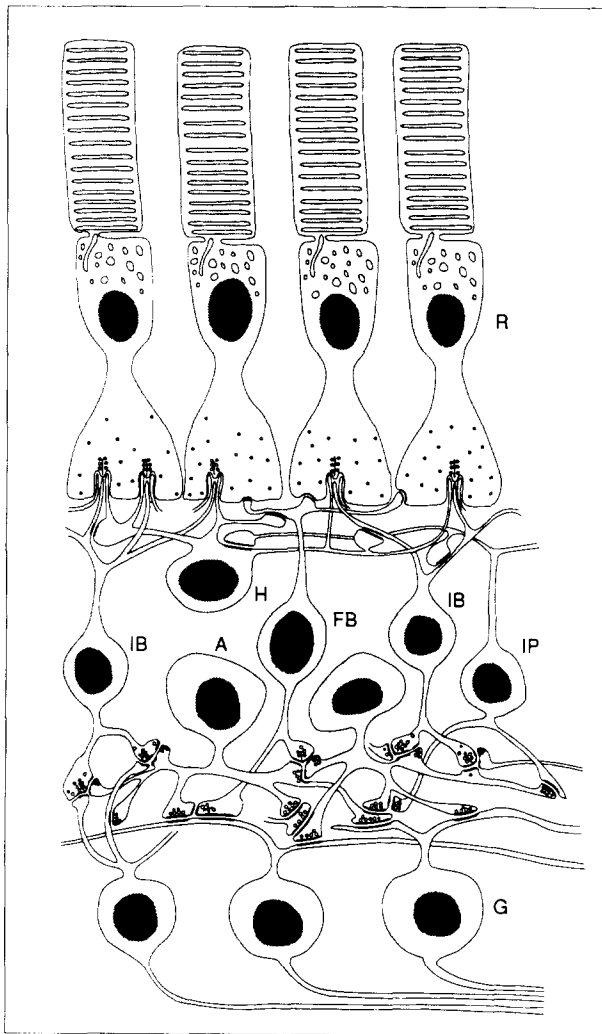


FIGURE 1. Cross section through the biological retina. R: photoreceptor, H: horizontal cell, IB: invaginating bipolar cell, FB: flat bipolar cell, A: amacrine cell, IP: interplexiform cell, G: ganglion cell. The outerplexiform layer is beneath the foot of the photoreceptors. The invagination into the foot of the photoreceptor is the site of the triad synapse. In the center of the invagination is a bipolar cell process, flanked by two horizontal cell processes.

describe our implementation of the model and compare its behavior with that observed in biological retinas.

PHOTORECEPTOR

The photoreceptor transduces an image focused on the retina into an electrical potential proportional to the logarithm of the local light intensity. The logarithmic nature of the response has two important system-level consequences:

1. An intensity range of many orders of magnitude is compressed into a manageable excursion in signal level.
2. The voltage difference between two points is proportional to the *contrast ratio* between the two corresponding points in the image, independent of incident light intensity.

The logarithmic nature of the output of the biological photoreceptor is supported by psychophysical and electrophysiological evidence. It is common experience that the perception of a scene does not change over a wide range of illumination levels. Psychophysical investigations of human visual sensitivity thresholds show that the threshold increment of illumination for detection of a stimulus is proportional to the background illumination over several orders of magnitude (Shapley, 1984). Physiological recordings from photoreceptors show that their electrical response to be logarithmic in light intensity over the central part of their range, as are the responses of other cells in the distal retina (Rodieck, 1973).

The primary transducer in our silicon retina is a photodetector described in (Mead, 1985). This photodetector is a vertical bipolar transistor, which occurs as a natural by-product in the CMOS process used for implementing the analog processing elements. This transistor produces approximately 100 electrons for every incident photon. The current from the phototransistor is fed into a circuit element with an exponential current-voltage characteristic, thereby creating an output voltage that is logarithmic in the incoming light intensity. The exponential element is realized by two diode-connected MOS transistors in series. In the sub-threshold range, corresponding to the current levels out of the phototransistor, the drain current of an MOS transistor is an exponential function of the gate-source voltage. We use two transistors to insure that the voltage range of the output is appropriate for subsequent pro-

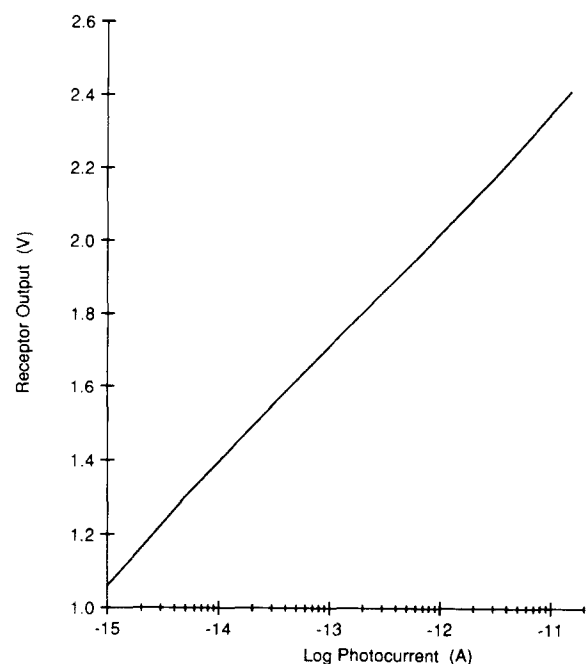


FIGURE 2. Measured response of logarithmic photodetector. Photocurrent is proportional to incident light intensity. Response is logarithmic over more than four orders of magnitude in intensity.

cessing by the kinds of amplifiers we can build in this technology. The voltage out of this photoreceptor is logarithmic over four or five orders of magnitude of incoming light intensity, as shown in Figure 2. The lowest photocurrent is about 10^{-14} amps, which translates to a light level of 10^5 photons per second. This level corresponds approximately to moonlight, which is about the lowest level of light visible using the cones in a vertebrate retina.

HORIZONTAL RESISTIVE LAYER

The horizontal cells in many species are connected to each other by gap junctions to form an electrically continuous network in which signals propagate by

electrotonic spread (Ehinger & Dowling, in press). The voltage at every point in the network thus represents a spatially weighted average of the photoreceptor inputs. The farther away an input is from a point in the network, the less weight it is given. The horizontal cells are usually modeled as passive cables, in which the weighting function decreases exponentially with distance.

Our silicon retina includes a hexagonal network of resistive elements, patterned after the horizontal cells of the retina. The network is constructed by linking each photoreceptor to its six neighbors with resistive elements, to form the hexagonal array shown in Figure 3. The CMOS technology does not include a resistor of sufficiently high value as an inherent part of the pro-

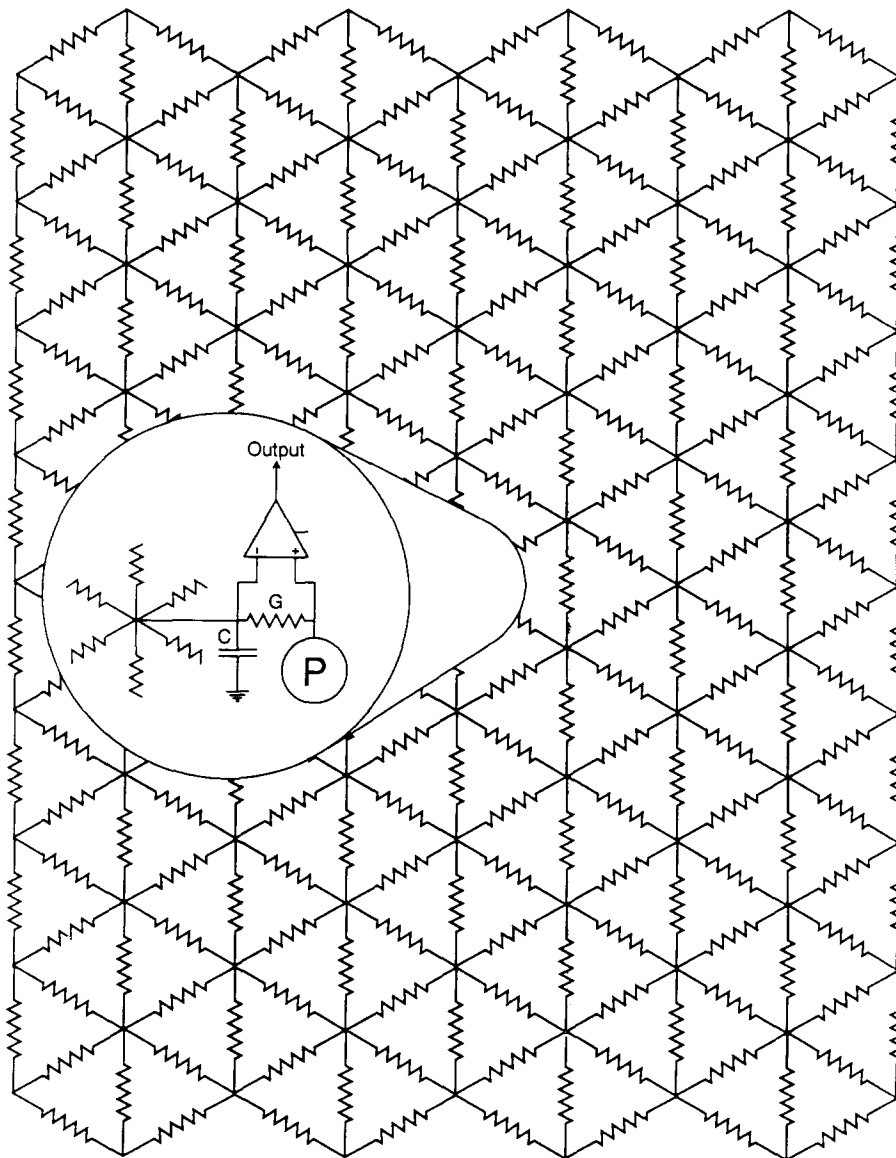


FIGURE 3. The silicon retina. Diagram of the resistive network and a single pixel element, shown in the circular window. The silicon model of the triad synapse consists of the conductance (G) by which the photoreceptor drives the resistive network, and the amplifier that takes the difference between the photoreceptor (P) output and the voltage on the resistive network. In addition to a triad synapse, each pixel contains six resistors and a capacitor C that represents the parasitic capacitance of the resistive network. These pixels are tiled in a hexagonal array. The resistive network results from a hexagonal tiling of pixels.

cess. All of our circuit components—resistors, capacitors, etc.—are made out of transistors (Sivilotti, 1987). Our resistive circuit has two advantages over a linear resistor.

1. The effective resistance of the connection can be controlled by an external input. This property is shared by the elements used in biological systems: The effective range of electrotonic spread in the horizontal cells is modulated, probably by dopamine released by the action of the inter-plexiform cells (Dowling, 1987).
2. The current-voltage relation of the element is linear for small voltage differences, but saturates at voltage differences larger than about 100 millivolts. Saturation is one of the more desirable properties of physical systems. In addition to warding off embarrassing infinities, saturation provides additional robustness to the collective system. For example, if one of our pixels fails and generates an output out of range, the damage it can do to the computation of the network is limited.

Both biological and silicon resistive networks have associated parasitic capacitances. The fine unmyelinated processes of the horizontal cells have a large surface-to-volume ratio, and hence their membrane capacitance to the extra-cellular fluid will average input signals over time as well as space. Our integrated resistive elements have an unavoidable capacitance to the silicon substrate, and hence provide the same kind of time-integration as their biological counterparts. The effects of delays due to electrotonic propagation in the network are most apparent when the input image is suddenly changed. Experiments in which this effect is dominant are discussed in the next section.

OUTER PLEXIFORM COMPUTATION

The receptive field of the bipolar cell shows an antagonistic center-surround response (Werblin, 1974). The center of the bipolar cell receptive field is driven by the photoreceptors, while the antagonistic surround is due to the horizontal cell influence. The triad synapse is thus the obvious anatomical substrate for this computation. In our model, the center-surround computation is a result of the interaction of the photoreceptors, the horizontal cells and the bipolar cells in the triad synapse.

One of the principle functions of this part of the biological retina is to prevent signals from saturating over the incredible dynamic range of the system. The first step in increasing the dynamic range is the logarithmic compression done by the photoreceptor. The next step is a level normalization, implemented by means of the resistive network. The horizontal cells of the retina provide a spatially averaged version of the receptor outputs, with which the local receptor potential can be compared. The triad synapse senses the differ-

ence between the receptor output and the potential of the horizontal cells, and generates a bipolar cell output from this difference. The maximum response occurs when the receptor potential is different from the space-time averaged outputs of many receptors in the local neighborhood. This situation occurs when the image is changing rapidly in either space or time.

The action of the horizontal cell layer is an example of lateral inhibition, a ubiquitous feature of peripheral sensory systems (von Békésy, 1967). Lateral inhibition is used to provide a reference value with which to compare the signal. This reference value is the operating point of the system. In the case of the retina, the operating point of the system is the local average of intensity as computed by the horizontal cells. Because it uses a local rather than global average, the eye is able to see detail in both the light and dark areas of high contrast scenes, a task that would overwhelm a television camera with only global adaptation.

The output of our silicon retina is analogous to the output of a bipolar cell in a vertebrate retina. Our triad synapse consists of two elements, as shown in Figure 3.

1. A conductance through which the resistive network is driven toward the receptor output potential.
2. An amplifier that senses the voltage difference across the conductance, and thereby generates an output proportional to the difference between the receptor output and the network potential at that location.

EXPERIMENTAL RESULTS

The voltage stored on the capacitance of the resistive network is the space and time averaged output of the photoreceptors, each of which contribute to the average with a weight that decreases with distance. Figure 4 shows the response of a single output to a sudden increase in incident illumination. Output from a real bipolar cell is provided for comparison. The initial peak represents the difference between the voltage at the photodetector caused by the step input and the old averaged voltage stored on the capacitance of the resistive network. As the resistive network equilibrates to the new input level, the output of the amplifier diminishes. The final plateau value is a function of the size of the stimulus, which changes the average value of the intensity of the image as computed by the resistive network.

Figure 5 shows the shift in operating point of the bipolar output of both a biological and a silicon retina as a function of surround illumination. Using the potential of the resistive network as a reference center, the range over which the output responds on the signal level averaged over the local surround. The full gain of the triad synapse can thus be used to report features of the image without fear that the output will be driven into saturation in the absence of local image information.

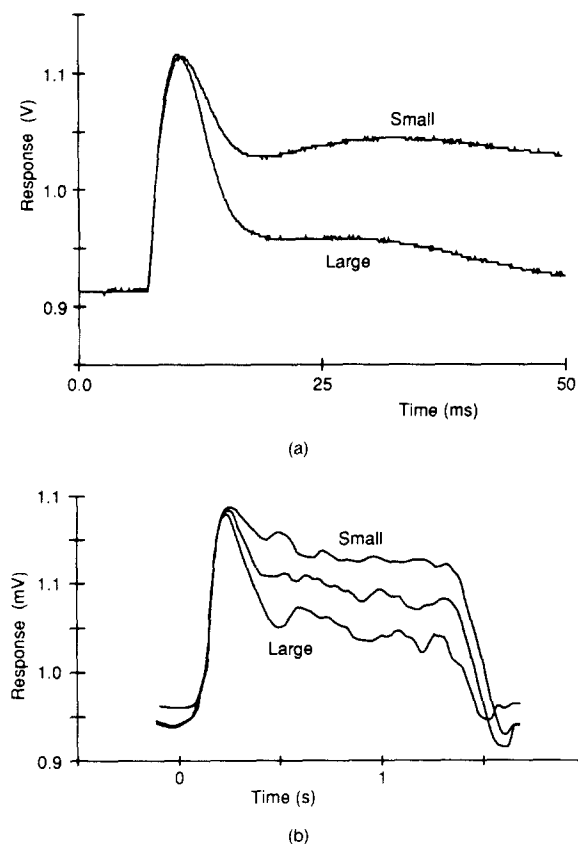


FIGURE 4. Temporal response of a bipolar cell of the mudpuppy, *Necturus maculosus*, and of a pixel in the silicon retina, to different size test flashes. Test flashes of the same intensity but different diameters were centered on the receptive field of the unit. (a) Response of a pixel. Larger flashes increased the excitation of the surround. The surround response was delayed due to the capacitance of the resistive network. Because the surround level is subtracted from the center response, the output shows a delayed decrease for long times. This decrease is larger for larger flashes. (b) Response of *Necturus* bipolar cell. Data from (Werblin, 1974).

The mechanisms evolved to keep the visual system operating over an enormous range of viewing conditions have important consequences with regard to the representation of data. In particular, the suppression of spatially and temporally smooth image information may be viewed as a filtering operation. The response of our silicon retina to a spatial intensity step is shown in Figure 6. The way the second spatial-derivative computation comes about is illustrated in Figure 7. A response of this type is produced by a receptive field that is a difference of Gaussians. A Laplacian filter, which has been used widely in computer vision systems, can be approximated by a difference of Gaussians (Marr, 1982). Both of these mathematical forms express, in an analytically tractable way, the computation that occurs as a natural result of an efficient physical implementation of local level normalization.

The output of the bipolar cells directly drive sustained type retinal ganglion cells of the mudpuppy,

Necturus maculosus. Consequently, the receptive field properties of this type of ganglion cell can be traced to the receptive field properties of the bipolar cells (Werblin, 1969). Although the formation of the receptive field of cat ganglion cells is somewhat more complex (Nelson, 1977), the end result is qualitatively similar. The response of a sustained type ganglion cell to a contrast edge placed at different positions relative to its receptive field is shown in Figure 6 (Enroth-Cugell, 1966). The spatial pattern of activity found in the cat is similar to that measured on our silicon retina.

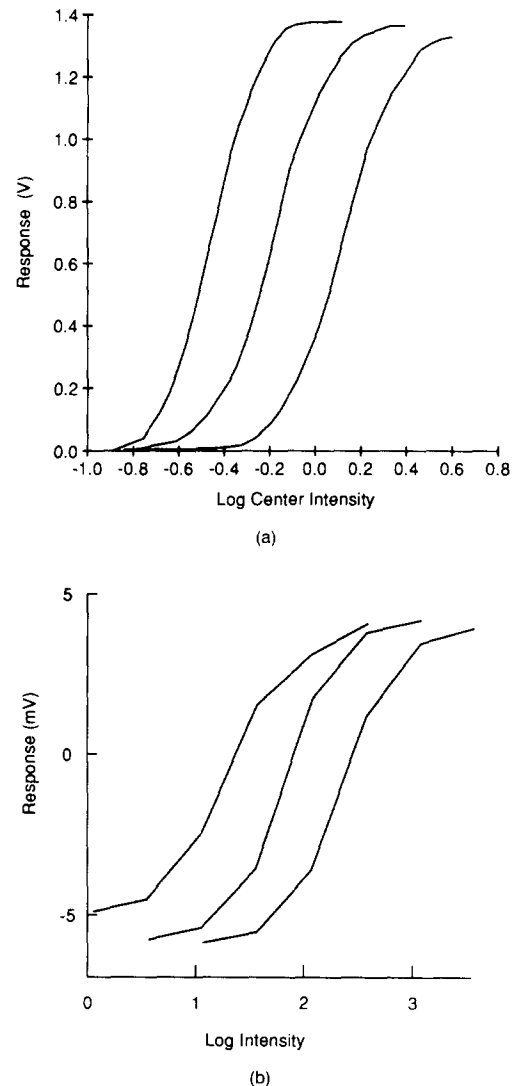
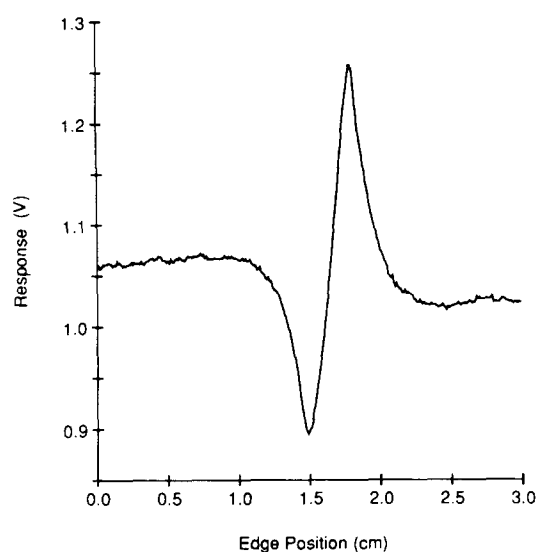
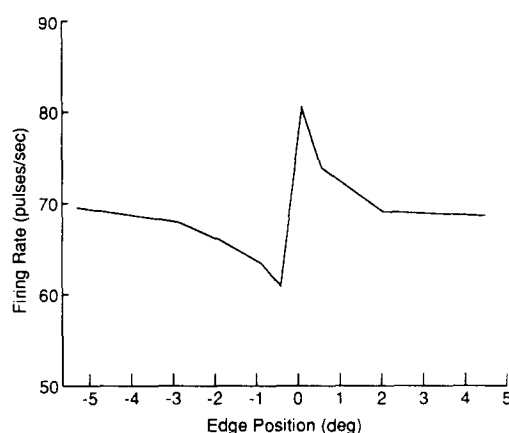


FIGURE 5. Curve shifting: Intensity-response curves shift to higher intensities at higher background illuminations. (a) Intensity response curves for a single pixel of the silicon retina. Curves plotted for three different background intensities. Stimulus was a small disk centered on the receptive field of the pixel. Steady state response was reported. (b) Intensity-response curves for a depolarizing bipolar cell elicited by full field flashes. Test flashes were substituted for constant background illuminations. These curves are plotted from the peaks of bipolar response to substituted test flashes. Peak responses are plotted, and measured from the membrane potential just prior to response. Data from (Werblin, 1974).



(a)



(b)

FIGURE 6. Spatial derivative response of a retinal ganglion cell and a pixel to a contrast edge. The vertical edge was held stationary at different distances from the receptive field center. Contrast of the edge was 0.2 log units in both experiments. (a) Pixel output measured at steady state as edge was moved in increments of 0.01 cm at the image plane. Interpixel spacing corresponded to 0.11 cm at the image plane. (b) On-center C-cell of the cat. The contrast edge was alternately turned on and off. The average pulse density over the period 10–20 seconds after the introduction of the edge was measured for each edge position (Enroth-Cugell, 1966).

DISCUSSION

The statement of the function of the retina is inseparable from the statement of its structure. Since Darwin's elucidation of the principle of natural selection, biological science has been able to state the function of organic structures. The function of the visual system is to see things about the world under a wide variety of illumination conditions thereby increasing the likelihood of that visual system being represented in the next generation. Unfortunately, this description of the function is not precise enough to design experiments

that can be conducted in the lifetime of an investigator. There is a great diversity in the theories that explain the purpose of the retina. Different investigators emphasize different aspects of retinal function such as spatial frequency filtering, adaptation and gain control, edge enhancement, or statistical optimization (Srinivasan, 1982). It is entirely in the nature of biological systems that the results of all the experiments designed to demonstrate one or another of these points of view can be explained by the properties of the single underlying structure. The evolved structure is able to subserve a multitude of purposes simultaneously.

We have taken the first step in simulating the computations done by the brain to process a visual image. We have used a medium whose structure is in many ways similar to neural structures. The constraints on

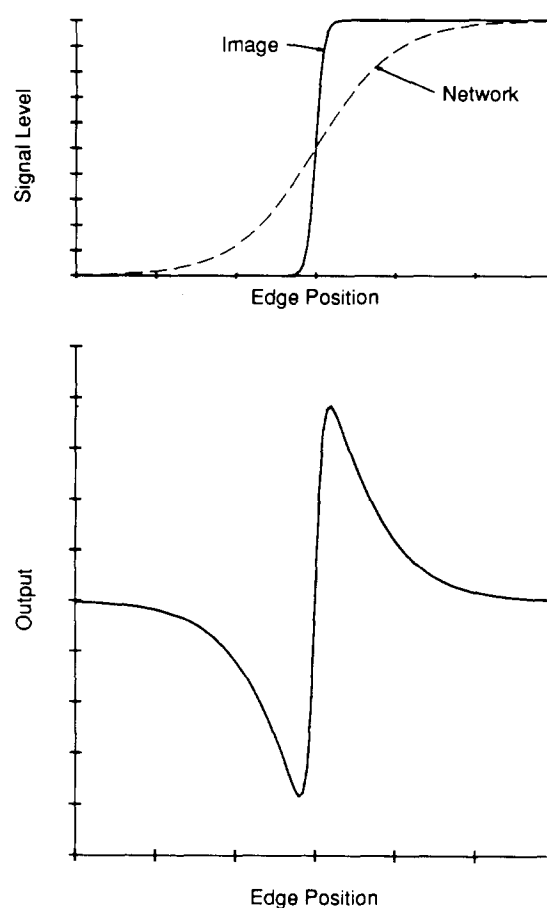


FIGURE 7. Model that explains the mechanism of the generation of pixel response to spatial edge in intensity. The solid line represents the voltage outputs of the photodetectors along a cross section perpendicular to the edge. The resistive network computes a weighted local average of the photoreceptor intensity, shown by the dashed line. The average intensity differs from the actual intensity at the stimulus edge because the photodetectors on one side of the edge pull the network on the other side towards their potential. The difference between the photodetector output and the resistive network is the predicted pixel output, shown in the lower trace. This mechanism results in increased output at places in the image where the first derivative of the intensity is changing.

our silicon systems are very similar to those on neural systems. The design is fairly compact; we can fit a 48×48 array of pixels on a chip that's one quarter of a square centimeter. As in the biological retina, density is limited by the wire length. The chip power efficient, using one hundred microwatts of power, and the computation is performed in real time.

In a small way, we have embarked upon a second evolutionary path—that of a silicon nervous system. As in any evolutionary endeavor, we must start at the beginning. Our first systems are simple and stupid. Compared to what the entire visual system of an animal does, or even what an actual retina does, our system is very low-level. It does, however, create a representation upon which higher level processing stages can be built; and that representation is true to its biological counterpart.

It is our conviction that our ability to realize simple neural functions is strictly limited by our understanding of their organizing principles, and not by difficulties in realization. If we *really* understand a system, we will be able to build it. Conversely, we can be sure that a system is not fully understood until a working model has been synthesized and successfully demonstrated.

The silicon medium can thus be seen to serve two complementary but inseparable roles:

1. To give computational neuroscience a synthetic element allowing hypotheses concerning neural organization to be tested.
2. To develop an engineering discipline by which real-time collective systems can be designed for specific computations.

The success of this venture will create a bridge between neurobiology and engineering, and will bring us

a much deeper view of computation as a physical process.

REFERENCES

- Dowling, J. (1987). *The retina: An approachable part of the brain*. Cambridge, MA: Harvard University Press.
- Enroth-Cugell, C., & Robson, J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, **187**, 517–552.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Mead, C. (1985, March). A sensitive electronic photoreceptor. In *1985 Chapel Hill Conference on Very Large Scale Integration* (pp. 463–471).
- Nelson, R. (1977). Cat cones have rod input: A comparison of the response properties of cones and horizontal cell bodies in the retina of the cat. *Journal of Comparative Neurology*, **172**, 109–136.
- Ratliff, F. (1965). *Mach bands: Quantitative studies on neural networks in the retina*. San Francisco: Holden-Day, Inc.
- Rodieck, R.W. (1973). *The vertebrate retina*. San Francisco: W.H. Freeman and Company.
- Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. In N. N. Osborne & G. J. Chader (Eds.), *Progress in retinal research* (Vol. 3, pp. 263–346). Oxford, England: Pergamon.
- Sivilotti, M.A., Mahowald, M.A., & Mead, C.A. (1987). Real-time visual computation using analog CMOS processing arrays. In *1987 Stanford Conference on Very Large Scale Integration*, Cambridge, MA: MIT Press.
- Srivivasan, M.V., Laughlin, S.B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London, Series B*, **216**, 427–459.
- von Békésy, G. (1967). *Sensory Inhibition*. Princeton, NJ: Princeton University Press.
- Werblin, F.S., & Dowling, J.E. (1969). Organization of the retina of the mudpuppy, *Necturus maculosus*. II. Intracellular recording. *Journal of Neurophysiology*, **32**, 339–355.
- Werblin, F.S. (1974). Control of retinal sensitivity II. Lateral interactions at the outer plexiform layer. *Journal of General Physiology*, **63**, 62–87.